"METHOD FOR DEVELOPING MEDICAL AND BIOCHEMICAL DIAGNOSTIC TESTS USING NEURAL NETWORKS," now abandoned. U.S. application Serial No. 08/798,306 is a continuation-in-part of U.S. application Serial No. 08/599,275. This application and U.S. application Serial No. 08/798,306 claim the benefit of priority under 35 U.S.C. §119(e) to U.S. provisional application Serial No. 60/011,449, entitled "METHOD AND APPARATUS FOR AIDING IN THE DIAGNOSIS OF ENDOMETRIOSIS USING A PLURALITY OF PARAMETERS SUITED FOR ANALYSIS THROUGH A NEURAL NETWORK" to Jerome Lapointe and Duane DeSieno, filed February 9, 1996.

**Please replace the paragraph beginning on page 6, line 27, through page 7, line 25, with the following paragraph:**

Because of the predictive and intuitive nature of medical diagnosis, attempts have been made to develop neural networks and other expert systems that aid in this process. The application of neural networks to medical diagnosis has been reported. For example, neural networks have been used to aid in the diagnosis of cardiovascular disorders (see, *e.g.*, Baxt (1991) "Use of an Artificial Neural Network for the Diagnosis of Myocardial Infarction," *Annals of Internal Medicine 115*:843; Baxt (1992) "Improving the Accuracy of an Artificial Neural Network Using Multiple Differently Trained Networks," *Neural Computation 4*:772; Baxt (1992), "Analysis of the clinical variables that drive decision in an artificial neural network trained to identify the presence of myocardial infarction," *Annals of Emergency Medicine 21*:1439; and Baxt (1994) "Complexity, chaos and human physiology: the justification for non-linear neural computational analysis," *Cancer Letters 77*:85). Other medical diagnostic applications include the use of neural networks for cancer diagnosis (see, *e.g.*, Maclin, *et al.* (1991) "Using Neural Networks to Diagnose Cancer" *Journal of Medical Systems 15*:11-9; Rogers, *et al.* (1994) "Artificial Neural Networks for Early Detection and Diagnosis of Cancer" *Cancer Letters 77*:79-83; Wilding, *et al.* (1994) "Application of Backpropogation Neural Networks to Diagnosis of Breast and Ovarian Cancer" *Cancer Letters 77*:145-53), neuromuscular

Q2 cont'd.

disorders (Pattichis, *et al.* (1995) "Neural Network Models in EMG Diagnosis", *IEEE Transactions on Biomedical Engineering 42*:5:486-495), and chronic fatigue syndrome (Solms, *et al.* (1996) "A Neural Network Diagnostic Tool for the Chronic Fatigue Syndrome", International Conference on Neural Networks, Paper No. 108). These methodologies, however, fail to address significant issues relating to the development of practical diagnostic tests for a wide range of conditions and do not address the selection of input variables.

**Please replace the paragraph on page 12, lines 9-21, with the following paragraph:**

Q3

The resulting systems are used as an aid in diagnosis. In addition, as the systems are used, patient data can be stored and then used to further train the systems and to develop systems that are adapted for a particular genetic population. This inputting of additional data into the system may be implemented automatically or done manually. By doing so the systems continually learn and adapt to the particular environment in which they are used. The resulting systems have numerous uses in addition to diagnosis, which includes assessing the severity of a disease or disorder, and predicting the outcome of a selected treatment protocol. The systems may also be used to assess the value of other data in a diagnostic procedure, such as biochemical test data and other such data, and to identify new tests that are useful for diagnosing a particular disease.

**Please replace the paragraph on page 13, lines 8-16, with the following paragraph:**

Q4

The methods are exemplified with reference to neural networks; however, it is understood that other data mining tools, such as expert systems, fuzzy logic, decision trees, and other statistical decision-support systems which are generally non-linear, may be used. Although the variables provided herein are intended to be used with decision-support systems, once the variables are identified, then a person, typically a physician, armed with knowledge of the

important variables can use them to aid in diagnosis in the absence of a decision-support system or using a less complex linear system of analysis.

**Please replace the paragraph on page 18, lines 16-20, with the following paragraph:**

Diagnostic software and exemplary neural networks that use the variables for diagnosis of endometriosis and the risk of delivery before a specified time are also provided. Software that generates a clinically useful endometriosis index is provided as software that generates an index for assessing the risk are provided.

**Please replace the paragraph on page 20, lines 22-23, with the following paragraph:**

FIGURES 16A and 16B show exemplary outputs from the software; FIGURE 16B includes the input data as well;

**Please replace the paragraph on page 24, lines 9-14, with the following paragraph:**

As used herein, partition means to select a portion of the data, such as 80%, and use it for training a neural net and to use the remaining portion as test data. Thus, the network is trained on all but one portion of the data. The process can then be repeated and a second network trained. The process is repeated until all partitions are used as test data and training data.

**Please replace the paragraph on page 26, lines 18-23, with the following paragraph:**

As used herein, performance of a system is said to be improved or higher when the results more accurately predict or determine a particular outcome. It is also to be understood that the performance of a system will typically be better as more training examples are used. Thus, the systems herein will improve over time as they are used and more patient data are accumulated and then added to the systems as training data.

**Please replace the paragraph beginning on page 27, line 23, through page 28, line 5, with the following paragraph:**

Q9

As shown herein, neural networks may also reveal that certain input factors that were not initially considered to be important can influence an outcome, as well as reveal that presumably important factors are not outcome determinative. The ability of neural networks to reveal the relevant and irrelevant input factors permit their use in guiding the design of a diagnostic test. As shown herein, neural networks, and other such data mining tools, are a valuable advance in diagnostics, providing an opportunity to increase the sensitivity and specificity of a diagnostic test. As shown herein, care must be taken to avoid the potential of poor-accuracy answer due to the phenomenon of local minima. The methods described herein provide a means to avoid this problem or at least minimize it.

**Please replace the paragraphs on page 31, lines 11-30, with the following paragraphs:**

**Isolation of important or relevant variables -ranking the variables**

Q10

Figures 3A-3B provide a flow chart of the process for isolating the important or relevant variables (Fig. 3A, Step E) within a diagnostic test. Such a process is typically conducted using a digital computer system to which potentially relevant information has been provided. This procedure ranks the variables in order of importance using two independent methods, then selects a subset of the available variables from the uppermost of the ranking. As noted above, other ranking methods can be used by those of skill in the art in place of chi square or sensitivity analysis. Also, if x is set to N (the total number of candidate variables), then ranking can be arbitrary.

The system trains a plurality of neural networks on the available data (Fig. 3A, Step I), as explained hereinafter, then generates a sensitivity analysis over all trained networks to determine to what extent each input variable was used in the network to perform the diagnosis (Fig. 3A, Step J). A consensus sensitivity analysis of each input variable is determined by averaging the individual

Q10 cont'd. sensitivity analysis results for each of the networks trained. Based upon sensitivity, a ranking order for each of the variables available from the patient history information is determined (Fig. 3A, Step K).

**Please replace the paragraphs beginning on page 35, line 14, through page 36, line 17, with the following paragraphs:**

**Chi-square contingency table**

Q11 When dealing with sparse binary data, a positive response on a given variable might be highly correlated to the condition being diagnosed, but occur so infrequently in the training data that the importance of the variable, as indicated by the neural network sensitivity analysis, might be very low. In order to catch these occurrences, the Chi-square contingency table is used as a secondary ranking process. A 2X2 contingency table Chi-square test on the binary variables, where each cell of the table is the observed frequency for the combination of the two variables (Fig. 3A, Step F) is performed. A 2X2 contingency table Chi-square test is performed on the continuous variables using optimal thresholds (which might be empirically-determined) (Fig. 3A, Step G)> The binary and continuous variables that have been based on Chi-square analysis are ranked (Fig. 3A, Step H).

The standard Chi-square 2X2 contingency table operative on the binary variables (Fig. 3A, Step F) is used to determine the significance of the relationship between a specific binary input variable and the desired output (as determined by comparing the training data with the known single output result). Variables that have a low Chi-square value are typically unrelated to the desired output.

For variables that have continuous values, a 2X2 contingency table can be constructed (Fig. 3A, Step G) by comparing the continuous variable to a threshold value. The threshold value is modified experimentally to yield the highest possible Chi-square value.

The Chi-square values of the continuous variables and of the binary variables can then be combined for common ranking (Fig. 3A, Step H). A

A11 contd.

second level of ranking can then be performed that combines the Chi-square-ranked variables with the sensitivity-analysis-ranked variables (Fig. 3A, Step L). This combining of rankings allows variables that are significantly related to the output but that are sparse (*i.e*, values that are positive or negative in only a small percentage of cases) to be included in the set of important variables. Otherwise, important information in such a non-linear system could easily be overlooked.

**Please replace the paragraph on page 37, lines 16-29, with the following paragraph:**

A12

In particular, the second level ranking process (Fig. 3A, Step L) starts by adding the highest ranked variable from the sensitivity analysis (Step K) to the set of important variables (Fig. 3A, Step H). Alternatively, the second level ranking process could be started with an empty set and then testing the top several (x) variables from each of the two sets of ranking. This second level ranking process uses the network training procedure (Fig. 3A, Step I) on a currently selected partition or subset of variables from the available data to train a set of neural networks. The ranking process is a network training procedure using the current set of "important" variables (which generally will initially be empty) plus the current variable being ranked or tested for ranking, and uses a greedy algorithm to optimize the set of input variables by myopically optimizing the input set based upon the previously identified important variable(s), to identify the remaining variable(s) which improve the output the most.

**Please replace the paragraphs beginning on page 38, line 9, through page 40, line 2, with the following paragraphs:**

A13

A number of other commercially available neural network computer programs may be used to perform any of the above operations, including Brainmaker™, which is available from California Scientific Software Co., Nevada Adaptive Solutions, Beaverton, OR; Neural Network Utility/2™, from NeuralWare, Inc., Pittsburgh, PA; NeuroShell™ and NeuroWindows™, from Ward Systems Group, Inc., Frederick, MD. Other types of data mining tools, *i.e.*, decision-

Q13

support systems, that will provide the function of variable selection and network optimization may be designed or other commercially available systems may be used. For example, NeuroGenetic Optimizer™ from BioComp Systems, Inc., Redmond, WA; and Neuro Forecaster/GENETICA, from New Wave Intelligent Business Systems (NIB5) Pte Ltd., Republic of Singapore, use genetic algorithms that are modelled on natural selection to eliminate poor-performing nodes within network population while passing on the best performing rates to offspring nodes to "grow" an optimized network and to eliminate input variables which do not contribute significantly to the outcome. Networks based on genetic algorithms use mutation to avoid trapping in local minima and use crossover processes to introduce new structures into the population.

Knowledge discovery in data (KDD) is another data mining tool, decision-support system, designed to identify significant relationship is that exist among variables, and are useful when there are many possible relationships. A number of KDD systems are commercially available including Darwin™, from Thinking Machines, Bedford, MA; Mineset™, from Silicon Graphics, Mountain View, CA, and Eikoplex™ from Ultragem Data Mining Company, San Francisco, CA. (Eikoplex™ has been used to provide classification rules for determining the probability of the presence of heart disease.) Others may be developed by those of skill in the art.

Proceeding with the ranking procedure, if, for example, x is set to 2, then the top two variables from each of the two ranking sets will be tested by the process (Fig. 3A, Steps L, S), and results are checked to see if the test results show improvement (Fig. 3B, Step T). If there is an improvement, the single best performing variable is added to the set of "important" variables, and then that variable is removed from the two rankings (Fig. 3B, Step U) for further testing (Fig 3A, Step S). If there is no improvement, then the process is repeated with the next x variables from each set until an improvement is found or all of the variables from the two sets have been tested. This process is repeated until either the source sets are empty, i.e., all relevant or important variables have

Q13

been included in the final network, or all of the remaining variables in the sets being tested are found to be below the performance of the current list of important variables. This process of elimination greatly reduces the number of subsets of the available variables which must be tested in order to determine the set of important variables. Even in the worst case, with ten available variables, the process would test only 34 subsets where $x = 2$ and only 19 subsets of the 1024 possible combinations if $x = 1$. Thus, where there are 100 available variables, only 394 subsets would be tested where $x = 2$. The variables from the network with the best test performance are thus identified for use (Fig. 3B, Step V).

**Please replace the paragraph on page 48, lines 3-17, with the following paragraph:**

**5. Method for evaluating the effectiveness of a diagnostic test course of treatment**

Q14

Typically, the effectiveness or usefulness of a diagnostic test is determined by comparing the diagnostic test result with the patient medical condition that is either known or suspected. A diagnostic test is considered to be of value if there is good correlation between the diagnostic test result and the patient medical condition; the better the correlation between the diagnostic test result and the patient medical condition, the higher the value placed on the effectiveness of the diagnostic test. In the absence of such a correlation, a diagnostic test is considered to be of lesser value. The systems provided herein, provide a means to assess the effectiveness of a biochemical test by determining whether the variable that corresponds to that test is an important selected variable. Any test that yields data that improves the performance of the system is identified.

**Please replace the paragraphs beginning on page 50, line 10, through page 51, line 24, with the following paragraphs:**

Q15

**Endometriosis**

The methods described herein, have provided a means to develop a non-invasive methodology for the diagnosis of endometriosis. In addition, the methods described herein provide means to develop biochemical tests that provide data indicative of endometriosis, and also to identify and develop new biochemical tests.

The methodology for variable selection and use of decision-support systems, has been applied to endometriosis. A decision-support system, in this instance, a consensus of neural networks, has been developed for the diagnosis of endometriosis. In the course of this development, which is detailed in the EXAMPLES, it was found that it was possible to develop neural networks capable of aiding in the diagnosis of endometriosis that only rely on patient historical data, *i.e.*, data that can be obtained from a patient by questionnaire format. It was found that biochemical test data could be used to enhance the performance of a particular network, but it was not essential to its value as a diagnostic tool. The variable selection protocol and neural nets provide a means to select sets of variables that can be inputted into the decision-support system to provide a means to diagnose endometriosis. While some of the identified variables include those that have traditionally been associated with endometriosis, others of the variables have not. In addition, as noted above, variables, such as pelvic pain and dysmenorrhea that have been associated with endometriosis are not linearly correlated with it to permit diagnosis.

Exemplary decision-support system are described in the Examples. For example, one neural net, designated pat07 herein, is described in Example 14. Comparison of the output of the pat07 network output with the probability of having endometriosis yields a positive correlation (see Table 1). The pat07 network can predict the likelihood of a woman having endometriosis based on her pat07 score. For example, if a woman has a pat07 score of 0.6, then she has a 90% probability of having endometriosis; if her pat07 score is 0.4, then she has a 10% probability of having endometriosis. The dynamic range of pat07 output when applied to the database was about 0.3 to about 0.7.

Q15 cont'd.

Theoretically, the output values can range from 0 to 1, but values below 0.3 or above 0.7 were not observed. Over 800 women have been evaluated using the pat07 network, and its performance can be summarized as follows:

**TABLE 1**

| Pat07 Score | Endometriosis (% of Total) |
|---|---|
| < 0.40 | 10 |
| 0.40 - 0.45 | 30 |
| 0.45 - 0.55 | 50 |
| 0.55 - 0.60 | 70 |
| > 0.60 | 90 |

**Please replace the paragraphs beginning on page 54, line 10, through page 55, line 11, with the following paragraphs:**

Q16

As described in the Examples, the pat07 is not the only network that is predictive of endometriosis. Other networks, designated pat08 through pat23a have been developed. These are also predictive of endometriosis. All these networks perform very similarly, and can readily be used in place of pat07. Thus, by following the methodology used to develop pat07, other similarly functioning neural nets can be and have been developed. Pat08 and pat09 are the most similar to pat07; these networks were developed by following the protocol outlined above, and were allowed to select important variables from the same set as that used for development of pat07.

It was found that the initial weighting of variables can have effects on the outcome of the variable selection procedure, but not in the ultimate diagnostic result. Pat08 and pat09 used the same database of patient data as pat07 to derive the disease relevant parameters. Pat10 through pat23a were training runs originally designed to elucidate the importance of certain parameters: history of endometriosis, history of pelvic surgery, dysmenorrhea and pelvic pain. For development of these, the importance of a variable was assessed by withholding that variable from the variable selection process. It was found that

Q16 in the variable selection process and training the final consensus networks, network performance did not significantly deteriorate.

Thus, although a particular variable, or set of variables, may have appeared to be significant in predicting endometriosis, networks trained in the absence of such variables do not have a markedly reduced ability to predict endometriosis. These results demonstrate (1) the effectiveness of the methodology for variable selection and consensus network training and (2) the adaptability of networks in general. In the absence of one type of data, the network found other variable(s) from which to extract that information. In the absence of one variable, the network selected different variables in its place and maintained performance.

**Please replace the paragraph on page 57, lines 18-28, with the following paragraph:**

Q17 Determination of impending birth is of importance, for example, for increasing neonatal survival of infants born before 34 weeks. The presence of fetal fibronectin in secretion samples from the vaginal cavity or the cervical canal from a pregnant patient after week 20 of pregnancy is associated with a risk of labor and delivery before 34 weeks. Methods and kits for screening for fetal fibronectin in body fluids and tissues, particularly in secretion samples from the vaginal cavity or the cervical canal, of a pregnant patient after week 20 of pregnancy are available (see, U.S. Patent Nos. 5,516,702, 5,468,619, and 5,281,522, and 5,096,830; see, also U.S. Patent Nos. 5,236,846, 5,223,440, 5,185,270).

**Please replace the paragraphs beginning on page 73, line 13, through page 74, line 29, with the following paragraphs:**

Q18 Table 1. Performance comparison of Qualitative fFN ELISA Test and the Preterm Delivery Risk Assessment Software Test relative to risk of preterm delivery before 35 completed weeks of gestation. The Risk Assessment Software combines fFN ELISA Test results with patient history and symptom

QB

information to provide a more accurately assessed risk of preterm delivery (before 35 completed weeks of gestation).

Table 2 compares the performance of the two tests relative to risk of preterm delivery within 7 days. The largest difference between the two tests is in the reduction of false positive test results of the software when compared to the ELISA test. The software decreased the number of false positive observations from 129 to 57, or about 56%. Accompanying this decrease in false positive results is the matching increase in true negative results, from 611 to 683. The true positive and false negative results remained essentially unchanged. The sensitivity and specificity of the software test is much higher than the ELISA test. Compare the sensitivity of 91.3% for the software with 87.0% for the ELISA, and the specificity of 92.3% for the software with 92.3% for the ELISA. Furthermore, the software test doubles the positive predictive value, increasing form 13.4% to 26.9%. Finally, the odds ratio is quadrupled and the relative risk more than tripled by the software.

| MEASURE | QUAL fFN ELISA TEST | RISK ASSESSMENT SOFTWARE TEST |
|---|---|---|
| True Positive | 20 | 21 |
| False Positive | 129 | 57 |
| True Negative | 611 | 683 |
| False Negative | 3 | 2 |
| Sensitivity | 87.0% | 91.3% |
| Specificity | 82.6% | 92.3% |
| Pos PV | 13.4% | 26.9% |
| Neg PV | 99.5% | 99.7% |
| Odds Ratio | 31.6 | 125.8 |
| Relative Risk | 27.4 | 89.7 |

Q18

Table 2. Performance comparison of Qualitative fFN ELISA Test and the Preterm Delivery Risk Assessment Software Test relative to risk of preterm delivery within 7 days.

Table 3 compares the performance of the two test relative to risk of preterm delivery within 14 days. Once again, the software decreases false positive test results when compared to the ELISA test, from 124 to 55, or about 53%. This decrease in false positive results is matched by the corresponding increase in true negative results, from 609 to 678. The number of true positive and false negative results were unchanged. While the sensitivity of the test was unaffected, the specificity of the test rose nearly 10 points, increasing from 83.1% to 92.5%. As seen before, the positive predictive value nearly doubled, increasing from 16.8% to 31.3%, and the odds ratio and relative risk increased substantially from 24.6 to 61.6 and from 20.7 to 44.7, respectively.

**Please replace the paragraph on page 77, lines 13-26, with the following paragraph:**

**Methodology Used**

Q19

The most commonly used method for determining the importance of variables is to train a neural network on the data with all the variables included. Using the trained network as the basis, a sensitivity analysis is performed on the network and the training data. For each training example the network is run in the forward mode (no training). The network outputs were recorded. Then for each input variable, the network is rerun with the variable replaced by its average value over the training example. The difference in output values is squared and accumulated. This process is repeated for each training example. The resulting sums are then normalized so that the sum of the normalized values equals the number of variables. In this way, if all variables contribute equally to the output, their normalized value would be 1.0. The normalized value can then be ranked in order of importance.

**Please replace the paragraphs on page 81, lines 13-26, with the following paragraphs:**

**Methodology Used**

Q20

When dealing with small training examples, the holdout method is effective in providing test information useful in determining network configuration and parameter settings. In order to maximize the data available for training without a big increase in processing time, a 20% holdout was used instead of the proposed 25%. This produced 5 partitions of the data instead of 4 and made 80% of the data for training in each partition.

To minimize the effects of the random starting weights, several networks were trained in the full training runs. In these runs three networks were trained in each of the five partitions of data, each from a different random start. The outputs of the networks were averaged to form a consensus result that has a lower variance than could be obtained from a single network.

**Please replace the paragraph on page 83, lines 17-31, with the following paragraph:**

**EXAMPLE 3**

**Preprocessing and input Western Blot Data Requirements**

Q21

The antigen data, from Western Blots, for the patients that was originally delivered to Logical Designs provided information on only the peak molecular weights and their associated intensities. Analysis of this data and of the original images from which the data was taken, suggests that it may be possible to use the original image digitized in a way that could provide more information to the neural network. In examining the original images for two experiments, preprocessing of the image data decreases the variability of the position of a specific molecular weight in the image. This preprocessing will use a polynomial fit through the standards image to produce a modified image. Preprocessing of the images will also include steps to normalize the background level and contrast of the images.

**Please replace the paragraph on page 85, lines 24-29, with the following paragraph:**

Q22 The weights of the network were then averaged together to generate a consensus value for each weight. Since the interconnection weight from the hidden element to the output could be either positive or negative, the weights were transformed so that all the output connections had the same sign. The weights were then averaged and the results plotted using Excel.

**Please replace the paragraph on page 86, lines 8-24, with the following paragraph:**

Q23 The plot of the sensitivity analysis and of the weights for the final consensus networks indicated that there are regions on the Western Blot that can aid in the prediction and diagnosis of the disease. The width of the regions of positive and negative correlation, as seen in the network weights, also indicates that the results shown are significant. If the peaks had been very narrow, one would have to conclude that the peaks were artifacts of the training process, similar to overtraining, and not from the underlying process being learned. The regions that appear important are as follows:

Positive Correlation:

31503.98 - 34452.12

62548.87 - 65735.97

84279.36 - 89458.49

Negative Correlation:

19165.9 - 20142.47

50263.36 - 53352.14

67725.77 - 78614.77.

**Please replace the paragraph on page 87, lines 4-18, with the following paragraph:**

Q24 **EXAMPLE 4**

**Investigate Fixed Input Dimension for Western Blot Data Requirements**

Q24 Cont.

Using peak molecular weights extracted from the preprocessed image, methods to reduce the varying dimension of the western blot data for a patient to a fixed dimension for the neural network will be investigated. This approach is desirable in that it will have substantially fewer network inputs than the full image approach. The basic problem is that the test yields a varying number of molecular weights that might by interrelated. Comparison of results from Example 5 and this example will indicate that patterns of molecular weights exist or if the weights are unrelated. Since there is some variability in the weight data, the approach to the processing of this data will be similar to a fuzzy membership function, even though the classification will be performed with a neural network.

**Please replace the paragraph beginning on page 88, line 22, through page 89, line 2, with the following paragraph:**

Q25

In order to test any of the data files produced by the above programs, the holdout methods was used with 80% of the data being used for training and the remaining 20% to be used for testing. Once the training data are produced form the Western Blot data, a random number column and the Patient ID column was added in the Excel spreadsheet. The data was then sorted on the random number column. This in effect shuffles the data. In this way, it is likely that each partition has examples from each of the gels. With these percentages, five separate training and test files were produced so as to allow a network performance to be estimated from the combined test set results.

**Please replace the paragraph beginning on page 101, line 23, through page 102, line 2, with the following paragraph:**

**Methodology Used**

Q26

Training examples were built for the Stage desired output and the AFS score desired output. There were 7 patients missing Stage information and 28 patients missing Score information. For the stage variable, the average value of 2.09 was used where the data were missing. For score, the missing data were replaced with a value depending on the value of the stage variable. For stage 1,

a score of 3 was used. From stage 2, 10.5 was used. For stage 3, 28 was used and for stage 4 the value 55 was used.

**Please replace the paragraph beginning on page 115, line 23, through page 116, line 6, with the following paragraph:**

**Conclusions**

The variable selection process appears to work well and has produced two alternative networks that work as well or better than the pat07 nets. The reason for this conclusion is that the performance statistics generated only on the training data appear slightly better for the pat07 than pat08 and pat09. Since the variable selection process carefully picks variables based on test set performance, the associated networks are not likely to have been overtrained. As a network becomes overtrained, the typical characteristic is that the training example performance increases, and the test set performance decreases. Thus, the higher performance of pat07 may be the result of slight overtraining.

**Please replace the paragraphs on page 124, lines 7-26, with the following paragraphs:**

**Results**

The count of variables found in the reduced subset run was smaller than that for the runs on the full training example. The typical performance of a consensus of networks was estimated using the holdout method with a partition of 5. The typical classification performance for the AFS run using the full training example was 77.22549%. The typical classification performance on the endo present subset was 63.008621%. If all examples were classified as negative, the performance for the full training example would be 78.82% and 65.29% for the subset. By changing the cutoff values for positive and negative classification better performance than suggested by these numbers can be achieved.

**Conclusions**

The results of the variable selection runs for the full training example and the subset of endo present examples suggest that the size of the training

Q28 contd.

example is of importance in the determination of the important variables. It is clear that as the size of the training example increases, more variables will be considered important. This result can also be interpreted as an indication that more training data will improve the variable selection process and also the overall performance of the consensus networks used in building the diagnostic test.

**Please replace the paragraph beginning on page 125, line 24, through page 129, line 10, with the following paragraph:**

**A. Variables**

Q29

The following are variables based on patient input data. Neural networks using all or selected subsets of these variables may be generated. Combinations of at least three of these variables may be used in conjunction with decision-support systems, particularly neural nets to predict risk of preterm delivery or impending delivery. The inputs for the variables are either yes, no, no answer, or a text input, such as age. The variables, listed by type are as follows:

1 Age

**Ethnic origin variables:**

2 EthOrg1: Caucasian;

3 EthOrg2: Black;

4 EthOrg3: Asian;

5 EthOrg4: Hispanic;

6 EthOrg5: Native American; and

7 EthOrg6: Other than the above.

**Marital status variables:**

8 MarSt1: Single;

9 MarSt2: Married;

10 MarSt3: Divorced/Separated;

11 MarSt4: Widowed;

12 MarSt5: Living with partner; or

13 MarSt6: Other than those listed above.

Gi29 cont'd.

**Education variables:**

14 Edu0: Unknown;

15 Edu1: <High School;

16 Edu2: High School Graduate; or

17 Edu3: College/trade.

**Patient complaint variables:**

18 PATIENT COMPLAINT #1 Patient has Uterine Contractions with or without pain;

19 PATIENT COMPLAINT #2 Patient has intermittent lower abdominal pain, dull, low backache, pelvic pressure;

20 PATIENT COMPLAINT #3 Patient has bleeding during the second or third trimester;

21 PATIENT COMPLAINT #4 Patient has menstrual-like or intestinal cramping;

22 PATIENT COMPLAINT #5 Patient has change in vaginal discharge or amount, color, or consistency; or

23 PATIENT COMPLAINT #6 Patient is not "feeling right".

**Variables from physician tests and assessments:**

24 Pooling refers to visual assessment to determine whether amniotic fluid has leaked into the vagina (see, *e.g.*, Chapter 36, Section 18, p. 657 in *Maternal Fetal Medicine: Principle and Practice*, 2nd Edition, Creasy, R.F. *et al.*, eds., W.B. Saunders & Co. (1989));

25 Ferning refers to the results of a test to detect the pattern formed when amniotic fluid is present in a cervical sample smeared on a clean slide and allowed to air dry (see, *e.g.*, Chapter 36, Section 18, p. 657 in *Maternal Fetal Medicine: Principle and Practice*, 2nd Edition, Creasy, R.F. *et al.*, eds., W.B. Saunders & Co. (1989));

26 Nitrazine refers to results from a known test used to measure the pH of amniotic fluid that has leaked into the vagina (see, *e.g.*, Chapter 36,

Cl 29 cont'd.

Section 18, p. 657 in *Maternal Fetal Medicine: Principle and Practice*, 2nd Edition, Creasy, R.F. *et al.*, eds., W.B. Saunders & Co. (1989));

27 estimated gestational based (EGA) on last period (LMP);

28 EGA by sonogram (SONO);

29 EGA by Best-EGA is the best of the EGA by SONO and EGA by LMP determined as follows:

if EGA by SONO is < 13 weeks, then EGA best is EGA SONO;

if the difference by EGA by LMP and EGA by SONO is > 2 weeks, then EGA best is EGA by SONO; otherwise EGA best is EGA by LMP;

30 EGA at Sampling refers to the EGA when fFN sampled;

31 CD INTERP, which refers to cervical dilatation (interpreted values - i.e. based on physicians estimates) where the number will be between 0 and 10 cm and is determined from the physicians response;

32 Gravity, which refers to the number of time woman has been pregnant;

33 Parity-term, which refers to the number of term births;

34 Parity-preterm, which refers to the number of preterm births;

35 Parity-abortions, which refers to the number of pregnancies ending in spontaneous or elective abortions;

36 Parity-living, which refers to the number of living children;

37 Sex within 24 hrs prior to sampling for fFN;

38 Vaginal bleeding at time of sampling;

39 Cervical consistency at time of sampling; and

40 UC INTERP, which refers to uterine contractions per hour as interpreted by the physician.

**Complications:**

Q29 cont.

41 0 COMP No previous pregnancies;

42 1 COMP have had at least one previous pregnancy without complications;

43 2nd comp at least one preterm delivery (delivery prior to 35 weeks);

44 3rd comp at least one previous pregnancy with a premature rupture of membrane (PROM);

45 4th comp at least one previous delivery with incompetent cervix;

46 5th COMP at least on previous pregnancy with pregnancy induced hypertension (PIH)/preeclampsia;

47 6th COMP at least one previous pregnancy with spontaneous abortion prior to 20 weeks;

48 OTHER COMP at least one previous pregnancy with a complication not listed above; and

49 RESULT - fFN ELISA qualitative test result (if positive value is 1, if negative value is 0).

---

**Please replace the paragraphs on page 129, lines 14-29, with the following paragraphs:**

---

Q30

**B. A first set of neural networks demonstrating that the methods herein can be used to predict pregnancy related events**

**EGA1-EGA4**

For these nets the preterm delivery is defined as less than or equal to 34 weeks, 0 days. The other nets herein (described below) define preterm delivery as less than or equal 34 weeks, 6 days.

Data were collected from the over 700 test patients involved in a clinical trial of the assay described in U.S. Patent No. 5,468,619. Variable selection was performed without fetal fibronectin (fFN) test data. The final networks,

Q30 cont'd

designated EGA1-EGA4, were trained with the variables set forth in the table below.

EGA1 - EGA4 represent neural networks used for variable selection. For EGA1, the variable selection protocol was performed on a network architecture with 8 inputs in the input layer, three processing elements in the hidden layer, and one output in the output layer. EGA2 is the same as EGA1, except that it is 9 inputs in the input layer.

**Please replace the paragraphs beginning on page 140, line 9, through page 141, line 5, with the following paragraphs:**

Q31

**D. Neural network prediction of risk of delivery within 7 days- EGAD7 and EGAD7F**

**1. Variable selection**

Using the same database described above for EGA1-EGA6, the variable selection protocol was applied to prediction of the risk for delivery within 7 days of sampling for the fFN test. As noted above for EGA5 and EGA6, the variable selection procedure was applied in the absence of the fFN test result. Application of the variable selection procedure to the 48 variables resulted in selection of the following variables:

1. Ethnic Origin 1: Caucasian (*i.e.*, yes or no);
2. Uterine contractions with or without pain (*i.e.*, yes or no);
3. Parity-abortions;
4. Vaginal bleeding at time of sampling;
5. Uterine contractions per hour;
6. No previous pregnancies.

**2. Neural nets**

Using these variables two consensus networks were trained. One, designated EGAD7, was trained without including the results of the fFN ELISA test result, and the other, designated EGAD7f, was trained with the results of the fFN ELISA test result.

**Please replace the paragraph on page 146, lines 4-8, with the following paragraph:**

Q32

EGAD7 is a set of 8 consensus networks trained similarly to EGAD7f, except that the input variables did not include the variable representing the result of the fFN ELISA test. This network can be used as a point of care application to give immediate results to the clinician rather than the 24 to 48 hours required to process the fFN sample.

**Please replace the paragraph beginning on page 153, line 1, through page 154, line 5, with the following paragraph:**

**EXAMPLE 14**

Q33

**Training of Consensus Neural Networks on Specific subsets of Pat07 Variables**

The example shows the results of a task designed to quantitate the contribution of pat07 variables to pat07 performance and to develop endometriosis networks using minimal numbers of pat07 variables.

Tasks:

1. Train final consensus networks using the following combination of Pat07 variables:
    a. All 14 minus Hx Endo (13 variables total)
    b. All 14 minus pelvic pain (13 variables total)
    c. All 14 minus dysmenorrhea (13 variables total)
    d. All 14 minus pelvic surgery (13 variables total)
2. Train final consensus networks using other combinations of Pat07 variables.
    a. Hx Endo, pelvic pain, and dysmenorrhea
    b. Hx Endo, pelvic pain, dysmenorrhea and Hx pelvic surgery
3. Train final consensus networks using other combinations of pat07 variables as indicated from above results.

**Please replace the paragraph beginning on page 156, line 23, through page 157, line 5, with the following paragraph:**

A34

Referring to Fig. 10, a final indicator pair C, D is based on an analysis of a consensus of preliminary indicator pairs from a plurality, specifically eight, trained neural networks 10A - 10H (Fig. 10). Each preliminary indicator pair A, B is provided to one of two consensus processors 150, 152 via paths 133-140 and 141-148. The first consensus processor 150 processes all positive indicators. The second consensus processor 152 processes all negative indicators. Each consensus processor 150, 152 is an averager, i.e., it merely forms a linear combination, such as an average, of the collection of like preliminary indicator pairs A, B. The resultant confidence indicator pair is the desired result, where the inputs are the set of clinical factors for the patient under test.

**IN THE ABSTRACT:**

Please amend the abstract as follows (a marked-up copy of the amended abstract is attached to this Amendment):

**Please replace the paragraph on page 168, lines 1-5, with the following paragraph:**

A35

**ABSTRACT**

Computer systems and methods for diagnosing endometriosis and for assessing the risk of delivery within selected time period after performing a test to assess the risk of preterm delivery or before thirty-five weeks of gestation are provided.

**IN THE CLAIMS:**

Please replace claim 5 with the following amended claim (a marked up copy of the amended claims is attached to this Amendment):

A36

5. (Amended) The system of claim 3, wherein the time period is within fourteen days of performing a biochemical test to measure fetal fibronectin.